

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

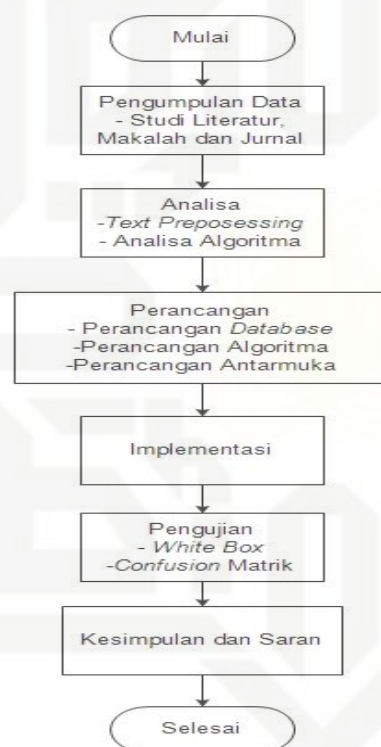
- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III

METODOLOGI PENELITIAN

Suatu penelitian diperlukan adanya metode pendekatan untuk melaksanakan suatu penelitian itu berjalan lancar dan sesuai dengan tujuan. Adapun tahapan pendekatan untuk penelitian dan menjadi panduan dalam menyelesaikan tugas akhir ini:



Gambar 3.1 Flowchart Metodologi Penelitian

Metodologi penelitian terdiri dari beberapa tahapan yang terkait secara sistematis. Tahapan ini diperlukan untuk memudahkan dalam penelitian. Tahapan yang dilakukan antara lain sebagai berikut:

3.1 Pengumpulan Data

Tahap pengumpulan data yaitu tahap mengumpulkan berbagai referensi yang diambil dari buku, jurnal, makalah, maupun *website*. Pengumpulan ini

bertujuan untuk membantu dalam mengumpulkan informasi yang berkaitan dengan penelitian yang dibuat. Untuk pengambilan data *tweet* melalui Twitter API (*Application Programming Interface*). Jumlah *tweet* yang akan dipakai adalah sebanyak 300 *tweet* berbahasa Indonesia yang diperoleh dari 5 akun meliputi akun perorangan yang biasa menggunakan kalimat tidak baku dan akun situs berita yang berisikan kalimat baku.

3.2 Analisa Permasalahan

Pada tahap ini dianalisa data yang digunakan dengan menggunakan cara mengumpulkan data terlebih dahulu, pengumpulan data dilakukan pada *tweet*, selanjutnya cara segmentasi kalimat, hapus *hashtag* dan URL, ekstraksi entitas, *cleaning*, tokenisasi kata, *stemming*, normalisasi kata, pos tagging, *phrase chunking* dan terakhir *sentence correction*

3.2.1 Analisa Data

Data yang diperoleh pada penelitian ini adalah *tweet*. Entitas pada *tweet* berupa teks, RT, *mention*, *hashtag*, URL, dan media.

3.2.2 Segmentasi Kalimat

Segmentasi kalimat adalah proses memecah teks menjadi kumpulan kalimat. Tokenisasi kalimat ini pemecahan dari RT, setelah itu pemecahan kalimat.

3.2.3 Hapus *Hashtag* dan URL

Untuk setiap kalimat yang sudah di segmentasi kalimat maka dihapus *hashtag* dan URL.

3.2.4 Ekstraksi Entitas

Ekstraksi entitas untuk mengidentifikasi pencocokan token sekuens di *tweet* ke dalam kamus. Dilakukan untuk mengidentifikasi kandidat untuk menemukan kandidat yang cenderung menjadi entitas. Entitas yang digunakan pada penelitian ini antara nya namaNegara, Provinsi, Aktor, Aktris, Atlet, Event Olahraga, Penyanyi, Suku, Pulau yang didapat dari DBPedia/Wikipedia.

3.2.5 Cleaning

Pembersihan karakter selain alfanumerik, tanda baca, karakter spasi dan emoji.

3.2.6 Tokenisasi Kata

Tokenisasi kata adalah proses pemecahan kalimat menjadi kata. Selain itu juga dilakukan pengubahan setiap huruf menjadi huruf kecil.

3.2.7 Stemming

Pengecekan kata menggunakan Algoritma *ECS (Enhanced Confix Stripping)* yang berdasarkan kata yang sudah ada di kata dasar yang di dapat dari Kateglo, apabila kata ada berarti kata tersebut Baku. Jika kata tersebut tidak Baku, maka dilakukan koreksi kata termasuk kata alay atau *typo* (Tahitoe Andita, 2010).

3.2.8 Normalisasi Kata

Untuk menemukan kata yang Baku dilakukan beberapa langkah yaitu pengkoreksian kata dengan menggunakan kamus alay, kamus singkatan dan fungsi *Levehnstein Distance*. Fungsi *Levehnstein* menghitung jarak kedekatan dari dua buah *string* melalui penambahan, penghilangan dan penggantian karakter hingga kedua *string* tersebut cocok (Muhammad Bahary Ilmy, 2006).

3.2.9 POS Tagging

Penentuan kelas kata (POS Tagging) adalah proses yang digunakan untuk memberikan *tagset* pada setiap kata. Penentuan kelas kata dilakukan untuk mempermudah mengetahui jenis kata setiap kata yang telah diproses sebelumnya. Kelas kata yaitu *verba*, *adjektiva*, *nomina*, *pronomina*, *adverbia*, *numeralia*, *interjeksi*, *konjungtor*, *preposisi* (Purwarianti, 2013). Algoritma Pos Tagging pada penelitian ini mengacu pada penelitian Ahmad Subhan Yazid (2017).

3.2.10 Phrase Chunking

Chunking adalah suatu proses mengklasifikasikan teks input untuk membentuk frasa yang dapat dikategorikan sebagai subjek (S), predikat (P), objek (O), Keterangan (K). Pendekatan yang disarankan untuk chunking adalah pendekatan

berbasis aturan. *Chunking* yang dilakukan pada penelitian ini adalah *Caps-Based Chunking* (*Chunking* berbasis huruf kapital).

3.2.11 Sentence Correction

Koreksi kalimat adalah tahapan dimana kalimat dikoreksi sesuai pola kalimat bahasa Indonesia. Koreksi dilakukan pada kalimat yang sudah melalui berbagai proses sebelumnya. Jika pola kalimat tidak sesuai, dilakukan dengan menempatkan unsur-unsur kalimat sesuai dengan pola yang memiliki unsur-unsur sama atau mendekati. Jika pola kalimat sudah sesuai maka proses selesai dan mendapatkan hasil yaitu kalimat Baku.

3.3 Perancangan

Perancangan digunakan untuk memudahkan dalam pembuatan aplikasi. Pada tahap ini aplikasi dirancang berdasarkan hasil analisa yang telah dilakukan pada proses sebelumnya dengan tujuan untuk lebih memudahkan dalam pembuatan aplikasi dan dapat mengetahui kebutuhan aplikasi serta mudah bagi user untuk menggunakannya. Tahapan dalam merancang aplikasi ini ada dua yaitu perancangan database dan perancangan antarmuka (*Interface*).

1. Tahapan Perancangan Basis Data atau Struktur *Database*

Setelah dilakukan analisa terhadap aplikasi yang dibuat, maka selanjutnya dilakukan perancangan basis data yang berisi tabel, *field*, dan atribut untuk melengkapi komponen aplikasi.

2. Tahapan Perancangan Antar Muka (*Interface*)

Agar terjalin hubungan baik antara pengguna dengan aplikasi, maka perlu dirancang antarmuka (*Interface*). Hal yang perlu ditekankan dalam perancangan antarmuka ini adalah bagaimana menciptakan tampilan yang menarik dan mudah dimengerti oleh pengguna (*user*) untuk aplikasi tersebut.

3.4 Implementasi

Pada proses implementasi pengembangan aplikasi ini dikembangkan pada spesifikasi perangkat keras dan perangkat lunak sebagai berikut:

Perangkat keras yang dibutuhkan antara lain:

<i>Processor</i>	: Core i3
<i>Memory</i>	: 2.00 GB

Perangkat Lunak yang dibutuhkan yaitu:

<i>Sistem operasi</i>	: Windows 7
<i>Bahasa Pemograman</i>	: PHP
<i>Tools Perancangan</i>	: Notepad ++
<i>Database</i>	: MySQL
<i>Web browser</i>	: UC Browser

3.4.1 Pengujian

Pengujian dilakukan dengan menggunakan *White Box* dan *Confusion Matrix*. *White Box* digunakan untuk pengujian aplikasi sedangkan *Confusion Matrix* untuk pengujian akurasi. Pengujian aplikasi ini bertujuan untuk melihat akurasi algoritma kalimat Baku yang dihasilkan, apakah berhasil atau tidak aplikasi tersebut. Pengujian dilakukan dengan kalimat tidak Baku hasil percakapan, misalnya pada media sosial Twitter

3.5 Penutup

Tahapan terakhir dari penelitian adalah kesimpulan berdasarkan hasil penelitian, serta akurasi yang didapat dari aplikasi pengubahan kalimat tidak Baku menjadi kalimat Baku. Serta memberikan saran dengan tujuan untuk menyempurnakan dan mengembangkan penelitian pada bidang yang bersangkutan.